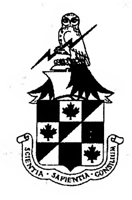
**SSC** Statistical Society of Canada
Société Statistique du Canada

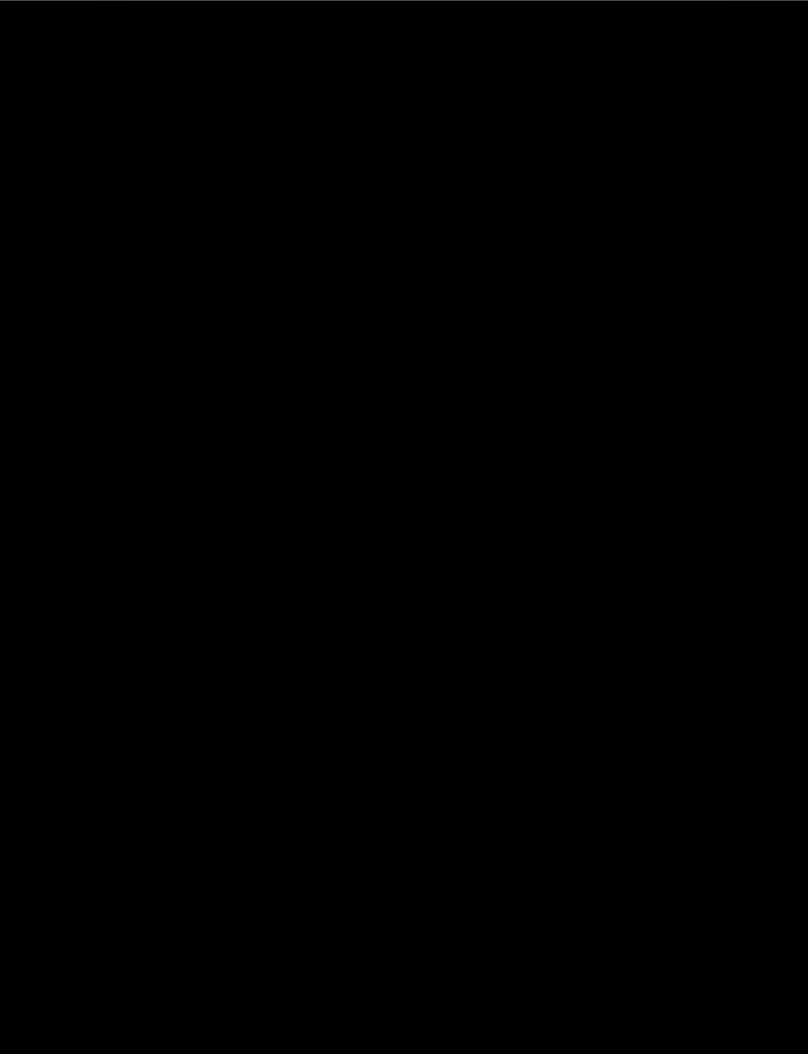
# 2000 Proceedings of the Survey Methods Section




# Recueil 2000 de la Section des méthodes d'enquête

**Papers presented at the 28th Annual Meeting of the Statistical Society of Canada, Ottawa, Ontario, Canada, June 4 - 7, 2000**

**Communications présentées au 28ième Congrès annuel de la Société statistique du Canada, Ottawa, Ontario, Canada du 4 au 7 juin 2000**



La section 2 présente les avantages et désavantages de cette méthode de collecte et fait un bref survol de la littérature à ce sujet. La section 3 décrit le Test du recensement national qui a constitué la source des données pour les analyses. Les sections 4 et 5 exposent les effets de la méthode sur la quantité et la qualité de l'information recueillie lors que l'enquête en analysant le cas de la variable portant sur le travail non rémunéré. Finalement, la section 6 présente les conclusions et recommandations.

## 2. DESCRIPTION ET REVUE DE LA LITTÉRATURE

Dans le cadre d'enquêtes menées auprès de ménages, le concept de réponses obtenues par procuration consiste à recueillir de l'information auprès d'un répondant à propos d'un ou de plusieurs membres de son ménage. Cette méthode présente de nombreux avantages, les principaux étant reliés à la simplification de la collecte des données. En théorie, si plus d'une personne est sélectionnée dans un ménage et si l'enquête se déroule à l'aide d'entrevues personnelles, une seule visite est nécessaire après avoir obtenu un rendez-vous avec le répondant. De plus, la durée d'entrevue par personne peut être raccourcie en raison de la non-nécessité de la part de l'intervieweur d'expliquer à nouveau les objectifs de l'enquête et le déroulement de l'entrevue. Cette méthode peut également augmenter la couverture de la population cible lorsque le seul moyen d'obtenir l'information désirée consiste à questionner un parent ou un proche. On peut, par exemple, penser aux enfants trop jeunes pour comprendre les concepts de l'enquête, aux personnes absentes pour toute la durée de la collecte ou aux gens qui présentent des problèmes de santé les empêchant de participer.

Il se peut cependant que le répondant ne soit pas la meilleure personne possible pouvant fournir l'information. Ainsi, à son insu, le représentant du ménage peut introduire des erreurs de réponse causées par sa méconnaissance du lien entre le concept de l'enquête et la personne cible. En outre, la tâche de cette personne est accrue d'autant plus que le nombre de personnes qu'elle représente est grand.

Une revue concise de la littérature au sujet des réponses obtenues par procuration se retrouve dans (Sudman, Bradburn et Schwarz, 1996). Ces auteurs présentent différents procédés utilisés pour mesurer les effets de cette méthode d'où émanent cinq conclusions générales. D'abord, les informations personnelles se voient accorder une plus grande importance à l'encodage (processus de mémorisation) et donc, lorsque questionné au sujet de ces informations, le processus de rappel est d'une plus grande qualité. On y conclut également que la distance sociale entre le répondant et la personne cible est un facteur très important dans la similitude des stratégies de réponse. Ce facteur est étroitement lié à la participation conjointe des deux individus dans l'événement ou l'objet de l'enquête et elle améliore également la qualité de l'information obtenue. On y définit aussi le biais de sélection comme étant la différence entre les caractéristiques des répondants et celles des membres de leur ménage. Ainsi, il se peut que systématiquement, une catégorie de gens soient plus susceptibles d'être les représentants de leur ménage. L'observation de différences entre les données des répondants et celles des membres de leur ménage ne montre pas obligatoirement un changement de comportement de la part des répondants lorsqu'ils fournissent des informations à propos de leurs proches, elles peuvent n'être que le résultat des différences intrinsèques des caractéristiques des ces deux groupes d'individus. Finalement, les répondants sont plus enclins à rapporter des événements socialement non souhaitables à propos d'autres personnes ce qui, dans ce cas, améliore la qualité des renseignements obtenus. À l'opposé, il se peut qu'ils méconnaissent le lien qui relie l'objet de l'enquête et la personne cible et, dans cette situation, qu'ils introduisent des erreurs de réponse.

## 3. LE TEST DU RECENSEMENT NATIONAL DE 1998

Le prochain recensement se tiendra le 15 mai 2001. Bien que le contenu de base du recensement de la population ait relativement peu changé depuis 1971, Statistique Canada continue d'apporter des modifications aux questions du recensement pour demeurer attentif aux besoins en information et pour ajuster le contenu des questions aux changements dans la structure socio-économique de la population canadienne.

Les deux questionnaires principaux utilisés lors du recensement de la population (2A et 2B) ont fait l'objet d'un test le 20 octobre 1998. Les objectifs de ce test furent d'améliorer le contenu des questionnaires (mise à l'essai de nouvelles questions, de nouvelles catégories de réponse et de changements quant à l'ordre d'apparition des questions) et d'évaluer une nouvelle méthode de collecte (contrôle centralisé). Le test était de nature volontaire et s'est étendu à 77000 ménages dans treize régions d'essai sélectionnées à travers le Canada. Tous les ménages faisant partie des secteurs de dénombrement formant ces régions ont reçu un questionnaire qu'ils ont dû retourner par la poste. Le choix de l'emplacement des régions d'essai

s'est appuyé sur des facteurs reliés à la fois aux opérations et au contenu et n'avait donc aucune composante aléatoire. Cependant, l'attribution des questionnaires à l'intérieur des régions d'essai s'est déroulée de façon systématique et donc, chacun des échantillons se voulait représentatif de chaque région. Un tiers des ménages a reçu le questionnaire 2A (questionnaire abrégé ne comportant que 6 questions), un autre tiers a reçu le questionnaire 2B.1 (questionnaire complet comportant environ 49 questions) et le dernier tiers a reçu une seconde version du questionnaire complet, le 2B.2. Les deux versions du questionnaire complet ont permis de mettre à l'essai de nouvelles questions. Naturellement, seules quelques-unes des modifications proposées seront apportées au questionnaire complet du recensement de 2001. Les taux de réponse obtenus furent de 80% pour le questionnaire 2A et de 70% pour les deux questionnaires complets.

Deux des trois traitements mis à l'essai lors du TRN (2B.1 et 2B.2) comportaient une question demandant au répondant d'identifier la ou les personnes ayant complété le questionnaire. Comme les questionnaires étaient de format matriciel, le répondant devait indiquer le numéro de la colonne où il avait disposé les informations le concernant. Il est très important de noter que ce type de question ne se retrouve habituellement pas dans le questionnaire du recensement de la population. Le tableau 1 donne la distribution des réponses obtenues à cette question pour les deux versions du questionnaire complet.

Il est important de noter que cette question permettait l'usage de réponses multiples. Ainsi, environ 80% des réponses multiples impliquaient les personnes 1 et 2. Dans ces cas, on peut supposer que les deux chefs du ménage se sont consultés pour compléter le questionnaire du test. Étant donné la présence de cette variable sur un questionnaire du recensement et de sa qualité relativement bonne (7% de non-réponse), une excellente occasion se présentait pour l'utiliser dans ce genre d'étude. Pour établir le profil des répondants et pour vérifier l'effet des réponses obtenues par procuration, seuls les ménages pour lesquels une seule personne (faisant partie du ménage) a répondu au questionnaire ont été conservés. Pour le questionnaire 2B.1, ceci représente 12894 questionnaires (72,7% des questionnaires reçus) et pour le questionnaire 2B.2, ceci représente 13092 questionnaires (74,8% des questionnaires reçus).

L'âge moyen des répondants au test s'est établi à 50,6 ans alors qu'il était de 46,3 ans pour l'ensemble des personnes couvertes par le test. La proportion de femmes ayant répondu au questionnaire fut de 58% alors qu'elles représentaient 52% des personnes couvertes par l'enquête.

## 4. EFFETS SUR LA QUANTITÉ D'INFORMATION OBTENUE

Une série de tests comparatifs a été élaborée pour tenter de déterminer si l'obtention des réponses par procuration a eu un effet sur la non-réponse, et ce, pour différentes variables de l'enquête. Le tableau 2 présente les taux de non-réponse (pour le répondant et pour les membres de son ménage) pour la variable portant sur l'état matrimonial en fonction du traitement et de l'âge du répondant.

Ainsi, on remarque que le taux de non-réponse est beaucoup plus faible pour le répondant que pour les membres de son ménage et ce, pour tous les groupes d'âge du répondant. Il est important de noter qu'historiquement les parents tendent à ne pas répondre à cette question pour les enfants de moins de 15 ans même si le questionnaire requiert une réponse à cette question pour eux. On pourrait alors croire que ceci constitue une explication aux taux de non-réponse plus élevés pour les autres membres du ménage.

**Tableau 1 : Distribution des réponses obtenues à la question : Qui a complété ce questionnaire ?**

| Réponses | Questionnaire 2B.1 Nombre | Questionnaire 2B.1 % | Questionnaire 2B.2 Nombre | Questionnaire 2B.2 % |
|---|---|---|---|---|
| Personne 1 | 9611 | 54,2 | 9697 | 55,4 |
| Personne 2 | 2981 | 16,8 | 3082 | 17,6 |
| Personne 3 | 219 | 1,2 | 221 | 1,3 |
| Personne 4 | 60 | 0,3 | 73 | 0,4 |
| Personne 5 | 19 | 0,1 | 14 | 0,1 |
| Personne 6 | 4 | 0,0 | 5 | 0,0 |
| Personne sur un autre questionnaire | 22 | 0,1 | 27 | 0,2 |
| Personne qui n'habite pas ici | 647 | 3,6 | 607 | 3,5 |
| Réponses multiples | 3003 | 16,9 | 2568 | 14,7 |
| Non-réponse | 1170 | 6,6 | 1197 | 6,8 |
| **Total** | **17736** | **100** | **17491** | **100** |

**Tableau 2 : Taux de non-réponse pour la variable portant sur l'état matrimonial en fonction de l'âge du répondant**

| Variable | Traitement | Âge du répondant | Taux de non-réponse pour le répondant | Taux de non-réponse pour les autres membres du ménage | Test de comparaison |
|---|---|---|---|---|---|
| État matrimonial | 2B.1 | 15-34 | 0,66 | 5,39 | Significatif |
| | | 35-44 | 0,56 | 3,89 | Significatif |
| | | 45-54 | 0,48 | 1,97 | Significatif |
| | | 55-64 | 0,42 | 2,03 | Significatif |
| | | 65 ans et plus | 0,56 | 3,41 | Significatif |
| | 2B.2 | 15-34 | 0,47 | 7,37 | Significatif |
| | | 35-44 | 0,23 | 6,06 | Significatif |
| | | 45-54 | 0,20 | 2,36 | Significatif |
| | | 55-64 | 0,71 | 2,39 | Significatif |
| | | 65 ans et plus | 0,75 | 5,24 | Significatif |

Le même exercice effectué en omettant les enfants de moins de 15 ans démontre des différences moins grandes entre les deux groupes, différences qui demeurent toutefois toujours significatives pour plusieurs groupe d'âge du répondant.

Des différences similaires ont été observées pour d'autres variables démographiques telles que l'union libre, la langue maternelle, la citoyenneté et le lieu de naissance des parents mais aussi pour des variables reliées à l'éducation et l'emploi. Aucune conclusion globale claire ne peut être émise à propos d'une relation entre l'âge du répondant et l'attitude qu'il adopte face à une question le concernant ou concernant un autre membre de son ménage, chacune des questions à l'étude montrant des tendances propres à elle-même.

La même série de test a été reprise en utilisant cette fois le sexe du répondant comme variable de classification. Le tableau 3 présente les résultats obtenus pour la variable portant sur le revenu personnel.

Encore une fois, on remarque que le taux de non-réponse est beaucoup plus faible pour le répondant que pour les membres de son ménage et ce, peu importe le sexe du répondant. Cette fois-ci, l'aspect relié aux enfants de moins de 15 ans ne peut expliquer une partie de ces différences puisque cette question ne s'applique pas à eux et qu'ils n'entrent pas dans le calcul du taux de non-réponse. Des différences significatives ont également été observées en utilisant cette classification pour plusieurs variables démographiques, reliées à l'emploi et à l'éducation. De la même façon que précédemment, chacune des variables doit être examinée pour établir un lien entre le sexe du répondant et son comportement et ce, en l'absence d'une tendance constante.

**Tableau 3 : Taux de non-réponse pour la variable portant sur le revenu personnel en fonction du sexe du répondant**

| Variable | Traitement | Sexe du répondant | Taux de non-réponse pour le répondant | Taux de non-réponse pour les autres membres du ménage | Test de comparaison |
|---|---|---|---|---|---|
| Revenu | 2B.1 | Femme | 3,22 | 5,37 | Significatif |
| | | Homme | 2,68 | 4,80 | Significatif |
| | 2B.2 | Femme | 3,44 | 5,20 | Significatif |
| | | Homme | 2,63 | 5,03 | Significatif |

## 5. EFFETS SUR LA QUALITÉ DE L'INFORMATION OBTENUE : CAS D'UNE QUESTION SUR LE TRAVAIL NON RÉMUNÉRÉ

La version 2B.1 du questionnaire complet mise à l'essai lors du TRN comportait une question relative au travail non rémunéré. On demandait alors au répondant d'indiquer le nombre d'heures consacrées aux activités à la maison pendant la semaine précédant le test. La première section de la question portait sur les travaux ménagers et l'entretien de la maison et du jardin, la deuxième section portait sur le temps consacré aux soins des enfants et la troisième section concernait le temps alloué pour offrir des soins ou de l'aide à une ou plusieurs personnes âgées et ce, toujours sans paye ou salaire. Pour chacune des sections, le répondant pouvait indiquer sa réponse à l'aide d'intervalles (0, moins de 5 heures, de 5 à 14 heures, de 15 à 29 heures, de 30 à 59 heures et 60 heures et plus)

Lors d'études qualitatives, il a été remarqué que cette question pouvait provoquer certaines discussions lorsque deux membres du ménage répondaient ensemble au questionnaire. Ainsi, l'interprétation de la question et l'évaluation du nombre d'heures qu'elle demandait pouvaient être propre à chaque personne. L'observation de ce comportement a motivé l'analyse des réponses obtenues par procuration pour cette variable (l'analyse a été restreinte à la première partie de la question, soit le nombre d'heures consacrées aux travaux ménagers). L'objectif était donc de démontrer l'existence d'un biais de représentation, c'est-à-dire une tendance qu'auraient les répondants à indiquer une réponse plus souhaitable, tout en s'assurant qu'il ne s'agissait pas d'un biais de sélection.

Dans un premier temps, l'échantillon a été divisé en deux sous-groupes. Il s'agit des femmes avec époux ou conjoint ayant elles-mêmes inscrit leur réponse (groupe A) et ensuite des femmes dont la réponse a été inscrite par leur époux ou conjoint (groupe B). En s'assurant de la présence de la caractéristique <avec époux ou conjoint> dans les deux sous-groupes, on augmente leur comparabilité et donc réduit le risque de biais de sélection. La stratégie de l'analyse fut la suivante. Toutes les femmes ont été classifiées dans un tableau à deux dimensions où différents groupes d'âge représentaient la première dimension (15-34, 35-44, 45-54, 55-64, 65 et plus) et où les intervalles d'heures de la question portant sur les travaux ménagers représentaient la deuxième dimension. Pour chacune des cellules de ce tableau, un test a ensuite été effectué pour déterminer si la proportion de femmes faisant partie des deux sous-groupes était significativement différente. Dans l'affirmative, on était soit en présence d'un biais de sélection ou bel et bien en présence d'un biais de représentation (ou d'une combinaison des deux). Pour le déterminer un modèle de régression logistique était développé où la variable dichotomique expliquée était l'appartenance au groupe A ou au groupe B et où les variables explicatives étaient des variables possiblement reliées à la question étudiée (travaux ménagers). Les variables explicatives retenues furent : une variable indicatrice relative à la détention d'un emploi salarié, le nombre d'enfants de moins de 15 ans, une variable indicatrice relative au travail à domicile, une échelle (1 à 10) du niveau du plus haut diplôme obtenu, la fréquentation d'un établissement scolaire, le mode de travail (temps plein ou temps partiel) et l'âge non regroupé. Si ces variables ne pouvaient pas expliquer la différence entre les proportions des femmes appartenant au groupe A et au groupe B dans chaque cellule (donc si l'existence du biais de sélection n'était pas démontrée) on concluait donc au biais de représentation.

Le tableau 4 présente les résultats des analyses. La mention <Groupe B> dans une cellule indique que la proportion de femmes appartenant à ce groupe est significativement plus grande que la proportion de femmes du groupe A, que les deux groupes sont comparables et que nous sommes probablement en présence d'un biais de représentation. La même règle a été employée pour les cellules portant la mention

**Tableau 4 : Détermination du biais de représentation pour les femmes - Variable traitant du travail non rémunéré**

| Âge | Catégories d'heures – Travaux ménagers | | | | | |
|---|---|---|---|---|---|---|
| | Aucune | Moins de 5 | 5 – 14 | 15 - 29 | 30 - 59 | 60 et plus |
| 15 - 34 | Groupe B | | | | | |
| 35 – 44 | | | | | | Groupe A |
| 45 – 54 | | | | | | |
| 55 – 64 | Groupe B | Groupe B | | | | |
| 65 et plus | Groupe B | | | Groupe B | Groupe A | Groupe A |

<Groupe A>. Les cellules vides indiquent que les proportions des deux groupes n'étaient pas significativement différentes ou si elles l'étaient, que les deux groupes étaient comparables en vertu des variables mentionnées auparavant.

La première observation importante découlant de l'examen du tableau résultant est que les différences significatives pour les catégories où le nombre d'heures consacrées aux travaux ménagers est élevé (30 à 59 heures, 60 heures et plus) sont le fruit d'une plus forte proportion de répondantes ayant indiqué ces catégories. Inversement, les différences significatives pour les catégories où le nombre d'heures est plus faible découlent d'une plus forte proportion de femmes du groupe B (réponses obtenues du conjoint). La majorité des différences significatives concernent des groupes comparables (aucune des variables du modèle de régression n'ayant été significatives) ce qui permet d'attribuer les différences observées à une stratégie de réponse distincte de la part des femmes ayant répondu pour elles-mêmes et des conjoints ayant indiqué une réponse pour leur épouse, donc à un biais de représentation.

Le même phénomène a été observé pour le tableau des hommes (les hommes ayant répondu pour eux-mêmes et les hommes pour lesquels la conjointe ou l'épouse a répondu à la question). Les répondants choisissent généralement les catégories reliées à un plus grand nombre d'heures et les conjointes choisissent des catégories reliées à un nombre d'heures plus faible. Cependant, une forte proportion des différences significatives concerne des groupes non comparables. Ainsi, il se peut que les différences observées ne proviennent uniquement que des différences de caractéristiques entre les répondants et les hommes pour lesquels leur conjointe a indiqué une réponse.

## 6. CONCLUSION ET RECOMMANDATIONS

L'effet des réponses obtenues par procuration lors du TRN a été mesuré de deux façons, soit à l'aide de la quantité d'information recueillie (taux de non-réponse) et de la qualité des réponses obtenues. Dans le cas de la quantité d'information, l'existence d'un tel effet est indéniable en raison d'un taux de non-réponse plus élevé chez les membres du ménage comparativement au répondant et ce, pour plusieurs variables à l'étude.

Pour ce qui est de la qualité des réponses obtenues, le cas de la variable concernant le travail non rémunéré a permis d'identifier des différences d'évaluation entre les répondants et les personnes pour lesquelles une réponse a été obtenue pour eux. Bien que non généralisé à tous les groupes d'âge et toutes les catégories de la variable, l'identification de ces différences sous le contrôle du biais de sélection permet de confirmer l'existence du biais de représentation. Étant donné la nature même de la question, aucune méthode ne permet de préférer la réponse par procuration ou du répondant lui-même, on ne peut que constater les différences.

Devant ces conclusions, il est clair que le fait de recueillir l'information à partir d'une personne représentant tous les autres membres de son ménage ne se réalise pas sans affecter la qualité des données. Évidemment, on ne peut négliger les nombreux avantages opérationnels de cette méthode. La réduction des coûts et la diminution du taux de non-réponse totale en sont les principaux. Les entrevues par procuration sont également nécessaires lorsqu'une composante importante de la population cible de l'enquête ne peut donner elle-même les informations la concernant (par exemple, dans une enquête sur les limitations d'activités, les personnes qui ont une incapacité trop sévère pour rendre possible une entrevue avec elles). Pour ces raisons, il est clair que cette méthode de collecte de l'information continuera d'être abondamment utilisé lors d'enquêtes menées auprès des ménages. Dans cette optique, il est important de tenir compte de la modification de la qualité des données lors des analyses. Ainsi l'ajout d'une question demandant d'identifier la personne ayant complété le questionnaire ou l'utilisation de données administratives pour valider les résultats peuvent être des moyens envisageables. On pourrait également développer un ensemble de questions de présélection qui augmenteraient les chances que le répondant soit vraiment le mieux placé pour fournir l'information désirée (diminution de la distance sociale).

## RÉFÉRENCES

Sudman, S., Bradburn, N.M., Schwarz, N. (1996). Thinking About Answers. *The Application of Cognitive Processes to Survey Methodology.* San Francisco : Jossey-Bass.

